Properties of Fisher information for Rician distributions and consequences in MRI

Jérôme Idier, Guylaine Collewet

Properties of Fisher information for Rician distributions and consequences in MRI

Jérôme Idier, *Member, IEEE,* Guylaine Collewet

Abstract

In magnetic resonance imaging, it is usual to perform several repeated acquisitions in order to reduce the impact of the noise on the evaluation of the average signal intensity. The main goal of this paper is to examine whether it is preferable to consider the magnitude of each complex measurement, or the single magnitude of the averaged data at each voxel. A thorough information theoretic study is proposed, which shows that the second option is preferable in all cases. In order to reach such a conclusion, several properties of the Fisher information of Rician distributions are proved, the main mathematical result being that it is an increasing function of the signal intensity, at constant noise level, when the latter is known. The case of an unknown noise level is also considered, with essentially the same conclusion. We also propose a more empirical comparison involving maximum likelihood estimation from simulated data, which is in agreement with the information theoretic analysis. Finally, an extension to imaging systems where multiple-receiver coils are employed is considered.

Index Terms

Magnetic resonance imaging (MRI), Rician distribution, Fisher information, Cramér-Rao lower bound, maximum likelihood estimation, multiple-receiver coils

I. Introduction

In the field of magnetic resonance imaging (MRI), the Gaussian law is a widely accepted model to describe the random character of the complex-valued data. However, incidental phase variations are observed in the acquired signal at different locations. Typically, all voxels in a homogeneous region share a common average signal intensity, whereas the assumption of a common average phase is not valid, because of inhomogeneities of the magnetic field. It is usually admitted that working with magnitude data allows to get rid of incidental phase variations. Since the magnitude of a complex-valued circular Gaussian random variable is known to follow a Rician law, magnetic resonance (MR) magnitude images are governed by such a distribution [1].

In a view to reduce the noise, it is common to acquire repeated measurements. N independent complex-valued observations $M_1, \ldots, M_N$ are then obtained for each voxel. Such a raw dataset will be called dataset (a). A usual preprocessing procedure consists in computing the so-called *magnitude data* as the magnitude of the averaged complex data $|M_1 + \cdots + M_N| /N$. It is understood here that no significant phase variation occurs through repetitions at a given voxel, which is a valid assumption according to our own practical experiments. Such magnitude data will be called dataset (b). It amounts to a single data point per voxel. An alternate preprocessing step would be to collect the magnitude data $|M_1|, \ldots, |M_N|$ separately. Up to the fact that N separate fast Fourier transforms are then needed (instead of a single one in the usual procedure), such a way to proceed is perfectly practicable. The corresponding magnitude data will be called dataset (c).

The primary goal of this paper is to determine which dataset is the most informative regarding the quantity of interest, *i.e.*, the magnitude A of the statistical mean of the data M_n. Indeed, we will mainly concentrate on a comparison between the inference from (b) versus (c), while many pros and cons of directly inferring from the complex data (a) can already be found in the literature (specifically, [2] is a reference paper in this respect). One important issue will be to evaluate the Fisher information related to A, and to compare the obtained expressions.

J. Idier is with the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN, UMR CNRS 6597), BP 92101, 1 rue de la Noë, 44321 Nantes Cedex 3, France. Tel: (+33)-2 40 37 69 09, e-mail: `Jerome.Idier@irccyn.ec-nantes.fr`.

G. Collewet is with IRSTEA, UR TERE, 17 avenue de cucillé, CS 64427, 35044 Rennes, France and wih Université Européenne de Bretagne, France. Tel: (+33)-2 23 48 21 67, e-mail: `Guylaine.Collewet@irstea.fr`.

Let us recall that the inverse of the Fisher information yields the CRLB, *i.e.*, a lower bound on the variance of any unbiased estimator. Asymptotically efficient estimators such as the maximum likelihood solution reach the CRLB asymptotically (*e.g.,* when the number of repeated measurements grows to infinity). However, far from the asymptote, the bias often plays an important role, especially at low signal to noise ratio. Therefore, additional comparisons will be performed to check that the performance of maximum likelihood estimation (MLE) in terms of mean-squared error (MSE) corroborates the analysis based on Fisher information, even when the maximum likelihood is significantly biased.

With a view to compare Fisher information values, an acknowledged difficulty is the lack of closed-form expressions for Rician-distributed data [2–6]. While previous contributions rely on numerical evaluations of the available integral expressions to analyze the informational content of magnitude datasets, the present paper proposes a mathematical study based on analytic properties of the Fisher information. Most of the obtained results are original and of stand-alone interest. Our main contribution is a monotony property of the Fisher information for the Rician distribution as a function of the so-called noncentrality parameter. We thoroughly study the case where the noise level is known, but we also consider the more difficult case where it is unknown. Moreover, we extend our main results to generalized Rician laws, which arise when multiple-receiver coils are employed and the resulting data are combined using the sum-of-squares method [7]. Concerning MLE, we put forward that the neg-log-likelihood is convex up to a simple change of variable, and we contribute to the practical computation of MLE by proposing a simple and rapidly converging Majorization-Minimization (MM) algorithm.

The rest of the paper is organized in five sections and an appendix. Section II provides a short mathematical reminder about the Rician probability, the Fisher information and the CRLB. Section III formulates the specific conclusions that can be drawn in the field of MRI from the proposed mathematical study, the latter being postponed to Section IV. A specific part of Section III considers magnitude estimation in each voxel using the MLE principle. Finally, Section V contains a concluding discussion. Such an organization allows different types of reading. Readers only interested by practical aspects of data acquisition and processing in the field of MRI could skip Section IV and the appendix, whereas readers primarily motivated by the mathematical side of the contribution can focus their attention on Section IV and on the accompanying appendix, which contain the most technical proofs.

II. The Fisher information of the Rician law related to its noncentrality parameter

A. Generalities

Let us define the Rician probability density (over $\mathbb{R}^+$) of parameters $\eta \geqslant 0$, $\alpha > 0$ as

$$R(x\,;\,\eta,\alpha) = \frac{x}{\alpha^2}\exp\left(-\frac{x^2+\eta^2}{2\alpha^2}\right) I_0\Big(\frac{\eta x}{\alpha^2}\Big)\,. \tag{1}$$

Here, η will be called the noncentrality parameter. A fundamental result from which the Rician law originates is as follows [8, p.100]. Let X_1, X_2 be independent, identically distributed (iid) centered real-valued Gaussian variables of standard deviation α, and let

$$M_1 = \eta\cos\phi + X_1, \quad M_2 = \eta\sin\phi + X_2, \tag{2}$$

where η and ϕ are two real constants, and $\eta \geqslant 0$. Then, $X = \sqrt{M_1^2+M_2^2}$ is a Rician random variable of parameters η, α, for any value of the angle parameter ϕ. In particular, it follows that the magnitude of MRI measurements acquired from single-coil is Rician distributed [1]. For the same reason, the Rician distribution also plays an important role in several other fields such as radar, sonar [9] and radio communications [10].

The Fisher information that an observable random variable X with probability density p carries about an unknown scalar parameter θ is defined by

$$\mathcal{I}_\theta = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log p(X;\theta)\right)^2\right].$$

The inverse of the Fisher information is the CRLB, *i.e.*, a lower bound for the variance of any unbiased estimator of parameter θ [11, chap. 12]. In the case of a multivariate parameter $\boldsymbol{\theta} = (\theta_j)_{j=1,\dots,J}$, the Fisher information becomes a positive semi-definite matrix:

$$\mathcal{I}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left(\frac{\partial \log p(X;\theta)}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log p(X;\theta)}{\partial \boldsymbol{\theta}}\right)^{\mathrm{t}}\right]$$

and the CRLB related to a given component θ_j is given by the jth diagonal entry of $\mathcal{I}_{\boldsymbol{\theta}}^{-1}$.

In the complex Gaussian case (2), the Fisher information related to η is easy to derive as

$$\mathcal{I}_\eta(M_1, M_2) = \frac{1}{\alpha^2}. \tag{3}$$

In the Rician case of $X = \sqrt{M_1^2 + M_2^2}$, several authors provide expressions of the Fisher information. A very complete study can be found in [3]. In the MRI context, the noncentrality parameter η is the parameter of interest, since it identifies with the image intensity, so we will only focus on the Fisher information related to η, akin to previous contributions such as [2, 5, 6]. Moreover, the two cases where α is known or unknown will be addressed.

B. Case of known parameter α

The Fisher information related to η when α is a known parameter reads [2, 3, 5, 6]:

$$\mathcal{I}_\eta = \int_0^\infty \frac{x^3}{\alpha^6} \exp\left(-\frac{x^2+\eta^2}{2\alpha^2}\right) \frac{I_1^2\left(\frac{\eta x}{\alpha^2}\right)}{I_0\left(\frac{\eta x}{\alpha^2}\right)} \,\mathrm{d}x - \frac{\eta^2}{\alpha^4}. \tag{4}$$

Unfortunately, the integral term does not admit a closed-form expression. However, the simple change of variable $y = \eta x/\alpha^2$ allows to reduce the remaining integral to a univariate function [3, 5, 12], since we have:

$$\mathcal{I}_\eta = \frac{\Psi(\rho)}{\alpha^2} \tag{5}$$

where $\rho = \eta^2/\alpha^2$ and

$$\Psi(\rho) = \int_0^\infty \frac{y^3}{\rho^2} \exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right) \frac{I_1^2(y)}{I_0(y)} \,\mathrm{d}y - \rho. \tag{6}$$

C. Case of unknown parameter α

1) Case of a single Rician variable: As found for instance in [2, 5, 6], the expression of the Fisher information matrix related to (η, α) reads:

$$\boldsymbol{I}_{\eta,\alpha} = \begin{pmatrix} \mathcal{I}_\eta & \mathcal{I}_{\eta,\alpha} \\ \mathcal{I}_{\eta,\alpha} & \mathcal{I}_\alpha \end{pmatrix}$$

where $\mathcal{I}_\eta$ is defined above,

$$\mathcal{I}_{\eta,\alpha} = \frac{2}{\alpha^2}\sqrt{\rho}\,(1 - \Psi(\rho)), \tag{7}$$

$$\mathcal{I}_\alpha = \frac{4}{\alpha^2}(\rho\Psi(\rho) - \rho + 1). \tag{8}$$

The determinant of $\boldsymbol{I}_{\eta,\alpha}$ is

$$\Delta(\rho) = \mathcal{I}_\eta\mathcal{I}_\alpha - \mathcal{I}_{\eta,\alpha}^2 = \frac{4}{\alpha^4}((\rho+1)\Psi(\rho) - \rho).$$

By explicit matrix inversion, one can deduce:

$$\boldsymbol{I}_{\eta,\alpha}^{-1} = \begin{pmatrix} \mathcal{J}_\eta & \mathcal{J}_{\eta,\alpha} \\ \mathcal{J}_{\eta,\alpha} & \mathcal{J}_\alpha \end{pmatrix} = \frac{1}{\Delta(\rho)} \begin{pmatrix} \mathcal{I}_\alpha & -\mathcal{I}_{\eta,\alpha} \\ -\mathcal{I}_{\eta,\alpha} & \mathcal{I}_\eta \end{pmatrix}$$

for all ρ such that $\Delta(\rho) > 0$. In particular, $\tilde{\mathcal{I}}_\eta = \mathcal{J}_\eta^{-1}$ gives the Fisher information related to the estimation of η when α is unknown:

$$\tilde{\mathcal{I}}_\eta = \frac{\Delta(\rho)}{\mathcal{I}_\alpha} = \frac{1}{\alpha^2}\tilde{\Psi}(\rho)$$

where

$$\tilde{\Psi}(\rho) = \Psi(\rho) - \frac{\rho(1-\Psi(\rho))^2}{\rho\Psi(\rho) - \rho + 1} \tag{9}$$

2) Case of several Rician variables: The previsous case is of limited applicability in the MRI context, since it either corresponds to a standard deviation that would be different for each voxel, or to an image (or a region) where the intensity would be known to be constant. In a more realistic context, we must rather consider a whole volume, where each voxel has its own average intensity, while the noise characteristics are common to all voxels [5]. In order to describe such a situation, let us assume that L independent random variables $\boldsymbol{X} = (X_1, \ldots, X_L)$ are observed, each X_ℓ being a Rician variable with parameters (η_ℓ, α), $\eta_\ell > 0$.

Because the laws of variables X_ℓ have the unknown parameter α in common, the estimation of the noncentrality parameters η_ℓ becomes a coupled problem. Let us denote $\tilde{\mathcal{I}}_{\eta_\ell}$ the Fisher information related to the estimation of η_ℓ when α and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)$ are unknown. The evaluation of $\tilde{\mathcal{I}}_{\eta_\ell}$ is substantially more complex than that of $\tilde{\mathcal{I}}_\eta$. However, the derivation follows the same lines, starting from the expression of the Fisher information matrix related to $(\boldsymbol{\eta}, \alpha)$:

$$\boldsymbol{I}_{\boldsymbol{\eta},\alpha} = \begin{pmatrix} \mathcal{I}_\eta(\rho_1) & \mathbf{0} & & \mathcal{I}_{\eta,\alpha}(\rho_1) \\ & \mathbf{0} \ \ddots & & \vdots \\ & & \mathcal{I}_\eta(\rho_L) & \mathcal{I}_{\eta,\alpha}(\rho_L) \\ \mathcal{I}_{\eta,\alpha}(\rho_1) & \ldots & \mathcal{I}_{\eta,\alpha}(\rho_L) & \sum_\ell \mathcal{I}_\alpha(\rho_\ell) \end{pmatrix}.$$

where $\mathcal{I}_\eta(\rho_\ell)$, $\mathcal{I}_{\eta,\alpha}(\rho_\ell)$, and $\mathcal{I}_\alpha(\rho_\ell)$ are respectively defined by (5), (7), and (8) at value $\rho = \rho_\ell = \eta_\ell^2/\alpha^2$. Let us remark that the $L \times L$ upper left corner of $\boldsymbol{I}_{\boldsymbol{\eta},\alpha}$ is diagonal. We can then deduce $\tilde{\mathcal{I}}_{\eta_\ell}$ as the reciprocal of the (ℓ, ℓ) entry of $\boldsymbol{I}_{\boldsymbol{\eta},\alpha}^{-1}$, using the partitioned matrix inversion lemma [13]. The resulting expression can be written as follows:

$$\tilde{\mathcal{I}}_{\eta_\ell} = \frac{1}{\alpha^2} \tilde{\Psi}_\ell(\boldsymbol{\rho}) \tag{10}$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_L)$ with $\rho_\ell = \eta_\ell^2/\alpha^2$, and

$$\tilde{\Psi}_\ell(\boldsymbol{\rho}) = \Psi(\rho_\ell) - \frac{\rho_\ell(1 - \Psi(\rho_\ell))^2}{D_\ell(\boldsymbol{\rho})} \tag{11}$$

with $D_\ell(\boldsymbol{\rho}) = \rho_\ell \Psi(\rho_\ell) - \rho_\ell + f_\ell(\boldsymbol{\rho})$ and

$$f_\ell(\boldsymbol{\rho}) = 1 + \sum_{i \neq \ell} \left(\rho_i + 1 - \frac{\rho_i}{\Psi(\rho_i)} \right). \tag{12}$$

The mathematical study of Section IV shows that $\lim_{\rho \to 0} \Psi(\rho)/\rho = \Psi'(0) = 1$, so that any voxel $i \neq \ell$ of null intensity has a vanishing contribution to $f_\ell(\boldsymbol{\rho})$. It also yields that $D_\ell(\boldsymbol{\rho}) > 0$ for all $\boldsymbol{\rho}$. Therefore, $\tilde{\Psi}_\ell(\boldsymbol{\rho})$ can be defined for all componentwise nonnegative $\boldsymbol{\rho}$.

A comparison between (11) and (9) indicates that $\tilde{\Psi}_\ell(\boldsymbol{\rho})$ and $\tilde{\Psi}(\rho_\ell)$ only differ by the presence of the additional partial sum in the expression of $f_\ell(\boldsymbol{\rho})$. As expected, the two quantities coincide when $L = 1$.

III. Application to MRI

For each voxel ℓ, let us assume that N repeated MR measures $M_{\ell 1}, \ldots, M_{\ell N}$ have been acquired, which form dataset (a) of iid complex-valued Gaussian variables of mean $A_\ell e^{i\phi_\ell}$ ($A_\ell \geqslant 0$) and common standard deviation σ. As a consequence,

(P1) $|M_{\ell 1} + \cdots + M_{\ell N}|/N$ (*i.e.*, dataset (b)) is a Rician variable of parameters $(A_\ell, \sigma/\sqrt{N})$.
(P2) $|M_{\ell 1}|, \ldots, |M_{\ell N}|$ (*i.e.*, dataset (c)) is a set of N iid Rician variables with parameters (A_ℓ, σ).

A. Estimation of the magnitudes A_ℓ when the noise level σ is known

In this case, there is no coupling between voxels, unless specific spatial regularity is accounted for, which is not considered here. Hence, for the sake of simplicity, we drop the voxel subscript ℓ in the present subsection, all the handled quantities being implicitly related to a generic voxel. Let $\mathcal{I}_A^{\mathrm{a}}$, $\mathcal{I}_A^{\mathrm{b}}$, and $\mathcal{I}_A^{\mathrm{c}}$ denote the Fisher information related to the dataset (a) and to the sets of transformed data (b) and (c), respectively.

On the one hand, we have

$$\mathcal{I}_A^{\mathrm{a}} = \frac{N}{\sigma^2} \tag{13}$$

according to (3). Let us remark that the single data point made of the sample mean $(M_1 + \cdots + M_N)/N$ would contain the same amount of information than dataset (a), since it is a sufficient statistics for the statistical mean in the Gaussian case.

On the other hand, (5), (P1), and (P2) yield

$$\mathcal{I}_A^{\mathrm{b}} = \frac{N}{\sigma^2}\,\Psi\!\left(\frac{NA^2}{\sigma^2}\right), \tag{14}$$

$$\mathcal{I}_A^{\mathrm{c}} = \frac{N}{\sigma^2}\,\Psi\!\left(\frac{A^2}{\sigma^2}\right). \tag{15}$$

According to expressions (13)-(15) and to general information theoretic principles, we can readily claim that:

- A smaller noise level σ increases the information contained in dataset (a).
- A larger number N of repetitions increases the information contained in datasets (a) and (c).
- The information contained in dataset (a) does not depend on the intensity value A, in contrast with the other two datasets.
- We have $\mathcal{I}_A^{\mathrm{a}} \geqslant \mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{a}} \geqslant \mathcal{I}_A^{\mathrm{c}}$, since the processing of a dataset cannot lead to an increase of the Fisher information, according to the data-processing inequality [14].
- Comparing the informational content of datasets (b) and (c) is equivalent to comparing two single magnitude data points at a constant noise level, the signal intensity being multiplied by a factor N in the case of (b).

On the other hand, it seems intuitive that $\mathcal{I}_A^{\mathrm{b}}$ does not decrease when N grows, but the mathematical validity of such an assertion clearly depends on the monotonicity of function Ψ. In the same way, a comparison between $\mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{c}}$ requires a better knowledge of the variations of Ψ. Theorem 2 of Section IV brings the missing piece, *i.e.*, that function Ψ is an increasing function. We can thus compare datasets (b) and (c) and draw some other conclusions, according to the following proposition.

Proposition 1 *For all $A > 0$, $\sigma > 0$ and $N > 1$,*

i) *the inequalities $\mathcal{I}_A^{\mathrm{a}} > \mathcal{I}_A^{\mathrm{b}} > \mathcal{I}_A^{\mathrm{c}}$ hold,*
ii) *$\mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{c}}$ are increasing functions of A, $1/\sigma$, and N,*
iii) *for arbitrary small values of A, $\mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{c}}$ vanish,*
iv) *for arbitrary large values of A, $\mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{c}}$ become equivalent to $\mathcal{I}_A^{\mathrm{a}}$.*

Let us remark that the last point corresponds to the convergence of the Rician distribution towards a Gaussian in the large magnitude regime [15]. On the other hand, the fact that $\mathcal{I}_A^{\mathrm{b}}$ and $\mathcal{I}_A^{\mathrm{c}}$ vanish for $A \to 0$ means that an unbiased estimator of A would have an unbounded variance when A becomes arbitrary small, as remarked in [16].

B. Estimation of the magnitudes A_ℓ when the noise level σ is unknown

In this case, there is a possible coupling between voxels through the estimation of σ, so we cannot examine each voxel separately. Let $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{a}}$, $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{b}}$, and $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{c}}$ denote the Fisher information related to A_ℓ when σ is unknown. On the one hand, we have $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{a}} = \mathcal{I}_{A_\ell}^{\mathrm{a}} = N/\sigma^2$ [2]. On the other hand, according to (10), (P1), and (P2), we have

$$\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{b}} = \frac{N}{\sigma^2}\,\tilde{\Psi}_\ell\!\left(\frac{N\boldsymbol{A}^2}{\sigma^2}\right) \tag{16}$$

$$\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{c}} = \frac{N}{\sigma^2}\,\tilde{\Psi}_\ell\!\left(\frac{\boldsymbol{A}^2}{\sigma^2}\right) \tag{17}$$

where $\tilde{\Psi}_\ell$ is defined by (11) and $\boldsymbol{A}^2 = (A_1^2, \ldots, A_L^2)$.

It is clear that $\tilde{\Psi}_\ell(\boldsymbol{\rho}) \leqslant \Psi(\rho_\ell)$ for all $\boldsymbol{\rho}$ since D_ℓ is positive, so that $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{b}} \leqslant \mathcal{I}_{A_\ell}^{\mathrm{b}}$ and $\tilde{\mathcal{I}}_{A_\ell}^{\mathrm{c}} \leqslant \mathcal{I}_{A_\ell}^{\mathrm{c}}$ (each inequality being strict for any strictly positive intensity). The differences $\mathcal{I}_{A_\ell}^{\mathrm{b}} - \tilde{\mathcal{I}}_{A_\ell}^{\mathrm{b}}$ and $\mathcal{I}_{A_\ell}^{\mathrm{c}} - \tilde{\mathcal{I}}_{A_\ell}^{\mathrm{c}}$ quantify the loss of information on A_ℓ when σ is unknown. It is also clear that any additional voxel contributing to the dataset increases the amount of information on the average intensities A_ℓ at all other voxels, since $f_\ell(\boldsymbol{\rho})$ will increase, and consequently, $\tilde{\Psi}_\ell(\boldsymbol{\rho})$ too. This is an intuitive result since any additional voxel will contribute to the estimation of σ, which is coupled to the estimation of A_ℓ for all ℓ.

A far less easy task is to establish a "tilded" counterpart of Proposition 1. Specifically, the main issue is to compare $\tilde{\mathcal{I}}^{\mathrm{b}}_{A_\ell}$ and $\tilde{\mathcal{I}}^{\mathrm{c}}_{A_\ell}$. Given (16)-(17), it relies on the variations of $\tilde{\Psi}_\ell(\mu\boldsymbol{\rho})$ as a function of a positive scalar μ, for any componentwise nonnegative vector $\boldsymbol{\rho}$. In Subsection IV-B, it is established that $\tilde{\Psi}_\ell(\mu\boldsymbol{\rho})$ is increasing with μ if the following inequality holds for all $\rho > 0$:

$$\rho\Psi'(\rho) > \Psi(\rho)(1-\Psi(\rho)). \tag{18}$$

The latter is numerically plausible, but it remains a conjecture.

The following proposition gathers the mathematical properties that can be deduced from Subsection IV-B.

Proposition 2 *The last two items of Proposition 1 hold when $\mathcal{I}^{\mathrm{a}}_A$, $\mathcal{I}^{\mathrm{a}}_A$, and $\mathcal{I}^{\mathrm{b}}_A$ are replaced by $\tilde{\mathcal{I}}^{\mathrm{a}}_{A_\ell}$, $\tilde{\mathcal{I}}^{\mathrm{b}}_{A_\ell}$, and $\tilde{\mathcal{I}}^{\mathrm{c}}_{A_\ell}$, for all $N > 1$, $L \geqslant 1$, $\boldsymbol{A} > 0$ (in the componentwise sense), and $\sigma > 0$. It is also true for the first two, with the proviso that* (18) *holds.*

C. Compared performance of MLE in terms of MSE

As already mentionned, dataset (a) cannot be less informative than datasets (b) and (c) according to the data-processing inequality [14]. In Sijbers and den Dekker's thorough study [2], such a general fact is corroborated by Figs. 1 and 4 therein. At the same time, an important and somewhat reversed finding is empirically obtained in [2], that MLE from magnitude data is preferable to MLE from complex data. Such a conclusion is obtained by comparing MSE values in practical situations, MLE from complex data being more strongly biased at low or moderate SNR. Let us also mention that an interesting study about bias reduction in the MRI context can be found in [16].

Akin to [2], we have found useful to complement our Fisher information study by an empirical evaluation of the statistical performance of MLE, in order to check whether it confirms the previous information based comparison between (b) and (c). It is the main goal of the present section. As a subsidiary contribution, we also propose an efficient iterative algorithm to maximize the likelihood of both datasets (b) and (c).

1) Maximization of the likelihood: Let us first consider the neg-loglikelihood (NLL) of dataset (c):

$$J(\boldsymbol{A},\sigma) = -\sum_{\ell,n} \log R(Y_{\ell n}; A_\ell, \sigma)$$

where $\boldsymbol{A} = (A_1, \ldots, A_L)$, and $Y_{\ell n}$ is a compact notation for $|M_{\ell n}|$. In a more explicit way,

$$J(\boldsymbol{A},\sigma) = \sum_\ell J_\ell(A_\ell,\sigma) + \log \sigma^{2LN} \tag{19}$$

with

$$J_\ell(A_\ell,\sigma) = \frac{NA_\ell^2}{2\sigma^2} + \sum_n \left(\frac{Y_{\ell n}^2}{2\sigma^2} - \log I_0\left(\frac{A_\ell Y_{\ell n}}{\sigma^2}\right) - \log Y_{\ell n} \right). \tag{20}$$

When the noise level σ is known, the maximum likelihood of $\boldsymbol{A}$ from dataset (c) is obtained as the minimizer $\hat{\boldsymbol{A}} = (\hat{A}_1, \ldots, \hat{A}_L)$ of J, where each $\hat{A}_\ell$ separately minimizes the corresponding component J_ℓ. When the noise level σ is unknown, joint minimization with respect to $\boldsymbol{A}$ and σ must be considered, which provides a couple $(\hat{\boldsymbol{A}}, \hat{\sigma})$. In all cases, the minimization of (19) has no closed-form solution, so it must be performed numerically using an iterative procedure. Several contributions have already addressed this problem, among which [3, 17]. However, none of them has highlighted the following strict convexity property.

Proposition 3 *The NLL function* (20) *is a strictly convex function of $\sqrt{A_\ell}$.*

Proof: The proof immediatly derives from the fact that $\log I_0(\sqrt{x})$ is a strictly concave function of x, or equivalently that the noncentral chi-square distribution with two degrees of freedom has a strictly log-concave density (provided that its noncentrality parameter is positive) [18]. ■

A consequence of Proposition 3 is that J_ℓ is a unimodal function of A_ℓ, *i.e.*, it admits a unique minimizer and no local minimum. The situation is more complex when J is considered as a function of $(\boldsymbol{A}, \sigma)$. Further studies would be needed to examine whether J is a unimodal function of $(\boldsymbol{A}, \sigma)$.

With a view to minimize J with respect to $\boldsymbol{A}$ and possibly to σ, another useful property is that $\log I_0(x)$ is a strictly convex function of x [19, 20]. Therefore, $-\log I_0(x)$ is majorized by its tangent, with equality at the considered point. As recently published in [21], such a property allows to build a fast and converging Majorization-Minimization (MM) algorithm to minimize J with respect to $\boldsymbol{A}$. MM schemes consist in iterately minimizing a majorizing approximation of the objective function. Such schemes intrinsically produce a non-increasing sequence of objective function values. For a more complete overview of the MM principle and its general convergence properties, the reader may consult [22].

When σ is known, it can be shown [17] that $A_\ell = 0$ is the minimizer of J_ℓ if and only if

$$\frac{1}{N}\sum_n Y_{\ell n}^2 \leqslant 2\sigma^2.$$

No iterative scheme is then required to minimize J_ℓ. Otherwise, the following MM iteration is an efficient way to converge to the minimizer:

$$A_{\ell k+1} = \frac{1}{N}\sum_n r_0\left(\frac{A_{\ell k} Y_n}{\sigma^2}\right) Y_n, \tag{21}$$

where $r_0(x)$ is the ratio between $I_1(x)$ and $I_0(x)$.

When σ is unknown, let us stress that the same MM principle remains applicable. From the majorization of $-\log I_0(x)$ by its tangent, we first deduce the following majorizing approximation of J at $(\boldsymbol{A}_k, \sigma_k)$:

$$K(\boldsymbol{A}, \sigma; \boldsymbol{A}_k, \sigma_k) = \sum_\ell K_\ell(A_\ell, \sigma; A_{\ell k}, \sigma_k) + \log \sigma^{2LN},$$

up to an additive constant, with

$$K_\ell(A_\ell, \sigma; A_{\ell k}, \sigma_k) = \frac{N A_\ell^2}{2\sigma^2} + \sum_n \left(\frac{Y_{\ell n}^2}{2\sigma^2} - r_0\left(\frac{A_{\ell k} Y_{\ell n}}{\sigma_k^2}\right)\frac{A_\ell Y_{\ell n}}{\sigma^2}\right).$$

It is then possible to obtain the joint minimizer of K with respect to $(\boldsymbol{A}, \sigma)$ under the following form:

$$\forall \ell, A_{\ell k+1} = \frac{1}{N}\sum_n r_0\left(\frac{A_{\ell k} Y_{\ell n}}{\sigma_k^2}\right) Y_{\ell n} \tag{22}$$

$$\sigma_{k+1}^2 = \frac{1}{2L}\sum_\ell \left(\frac{1}{N}\sum_n Y_{\ell n}^2 - A_{\ell k+1}^2\right) \tag{23}$$

which constitutes a joint MM scheme to minimize J with respect to $(\boldsymbol{A}, \sigma)$. Let us mention that a comparable algorithm has been derived in the Expectation-Maximization (EM) framework for fMRI data processing in [23].

In practice, the initial point $(\boldsymbol{A}_0, \sigma_0)$ must be chosen in the positive orthant, given the following proposition.

Proposition 4 *Let us assume that for each voxel ℓ, at least one data point $Y_{\ell n}$ is positive. Then the whole sequences $(A_{\ell k})$ and (σ_k) are strictly positive provided $A_{\ell 0} > 0$ for all ℓ and $\sigma_0 > 0$. Conversely, for any ℓ, k, $A_{\ell k} = 0$ implies $A_{\ell k+1} = 0$.*

Proof: Let us assume $A_{\ell k} > 0$ for all ℓ and $\sigma_k > 0$. Function $r_0(x)$ is strictly increasing from $r_0(0) = 0$ to $r_0(\infty) = 1$ [24]. Thus, $0 \leqslant A_{\ell k+1} \leqslant m_\ell$ for all ℓ, where m_ℓ denotes the empirical mean $m_\ell = \frac{1}{N}\sum_n Y_{\ell n}$, and the two bounds are strict unless $Y_{\ell n}$ cancels for all n. Moreover,

$$\sigma_{k+1}^2 \geqslant \frac{1}{2L}\sum_\ell \left(\frac{1}{N}\sum_n Y_{\ell n}^2 - m_\ell^2\right) \geqslant 0,$$

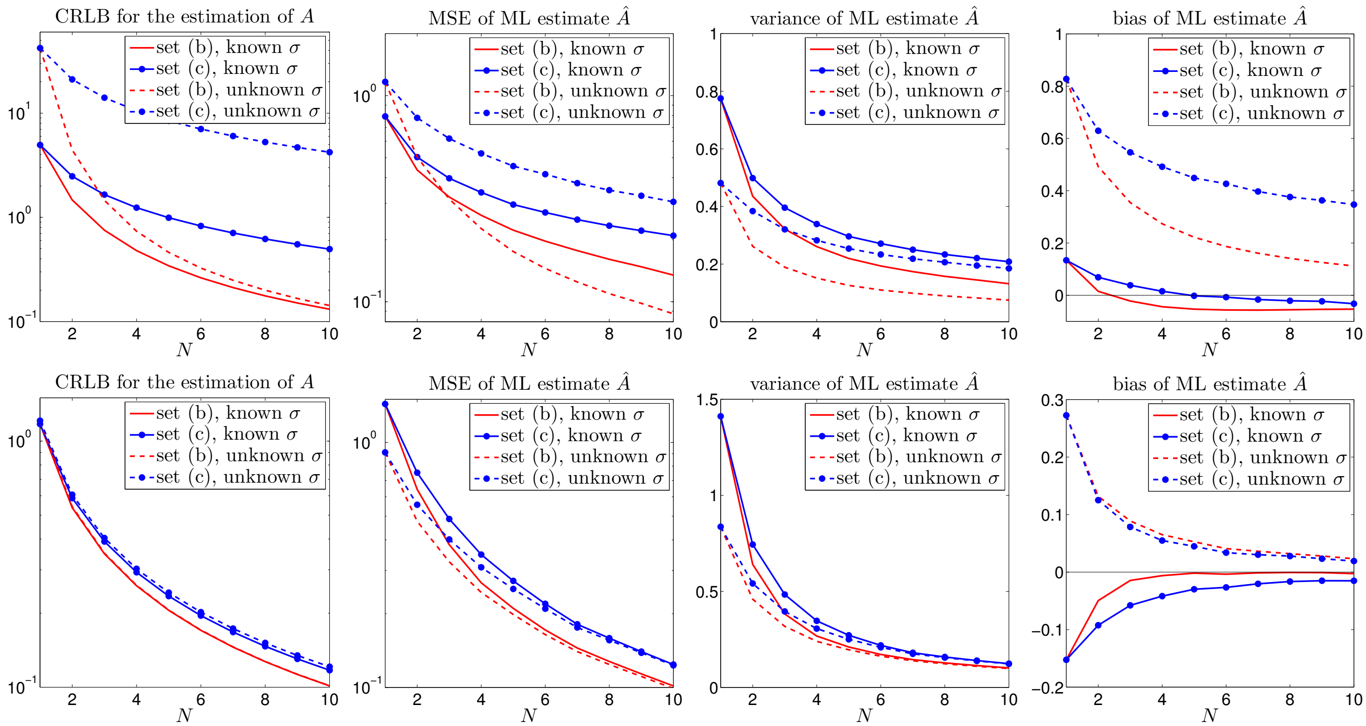

Fig. 1. Upper line: theoretical CRLB and MSE, variance, and bias of maximum likelihood for sets (b) and (c) in the cases when σ is known or unknown ($L = 10$) for $A = 0.5$ and $\sigma = 1$. Lower line: same for $A = 2$

the first inequality being strict if there is at least one couple (ℓ, n) for which $Y_{\ell n} > 0$. Finally, the last implication holds since $r_0(0) = 0$. ■

A rigorous mathematical convergence study of (22)-(23) remains to be made. General properties of MM schemes ensure that $(J(A_{\ell k}, \sigma_k))$ is a nonincreasing sequence. Based on [25], it is reasonable to assume that (22)-(23) converges towards a stationary point of J.

So far, we have designed a minimization scheme to compute the MLE solution in the case of dataset (c). Since it corresponds to observing a single Rician variable of parameters $(A_\ell, \sigma/\sqrt{N})$ at each voxel, dataset (b) can be processed using the same MM algorithm with a single data point for each voxel, up to a trivial multiplicative correction (by a factor $\sqrt{N}$) on the second parameter. However, it happens that the MLE solution becomes trivial, only in the case where σ is unknown. Indeed, it can be shown mathematically that the likelihood is not upper bounded in case (b), and that it becomes infinite at the MLE solution defined by $\hat{\sigma} = 0$, and $\hat{A}_\ell = |M_{\ell 1} + \cdots + M_{\ell N}| / N$ for all ℓ. Such a behavior would not occur if we considered regions of constant intensity, akin to [2, Eq. (46)]. Another possibility would be to add a noninformative penalization term to the likelihood in order to obtain a bounded criterion. In both cases, the proposed MM scheme (22)-MM2b would remain valid, up to simple adaptations. Anyway, the obtained intensity estimator is not a pathological one, so we proceed with it in the next subsection.

2) Statistical performance of MLE: We have simulated simple MRI data using the Rician model with noncentrality parameter values $A = 0.5$ and $A = 2$, and noise standard deviation $\sigma = 1$. In both cases, we have evaluated the theoretical CRLB, *i.e.*, $\mathcal{I}_A^{-1}$ or $\tilde{\mathcal{I}}_A^{-1}$, whether σ is assumed known or not. In the latter case, a region of $L = 10$ voxels of equal intensity has been considered. We have applied the proposed MM algorithms to repeatedly simulated data to evaluate the MLE solutions and to deduce empirical evaluations of the MSE, variance, and bias, as functions of the number N of repetitions, between 1 and 10.

Fig. 1 displays the results for $A = 0.5$ and $A = 2$. When $N = 1$, the two datasets (b) and (c) coincide, so the couples of corresponding curves coincide as well. For all $N > 1$, dataset (b) yields a lower value of CRLB than dataset (c) (which illustrates that $\mathcal{I}_A^{\mathrm{b}} > \mathcal{I}_A^{\mathrm{c}}$ and $\tilde{\mathcal{I}}_A^{\mathrm{b}} > \tilde{\mathcal{I}}_A^{\mathrm{c}}$). The difference is less sensible for larger intensity values, which is easily explained by the last point of Proposition 1 and by its equivalent in Proposition 2.

The main information given by Fig. 1 is that MLE statistically performs better from dataset (b) than from dataset

(c) in terms of MSE. Let us stress that neither the variance nor the MSE is directly comparable to the CRLB, because of the effect of the bias at low signal-to-noise ratio. As already pointed out and commented in [16], the MSE becomes smaller than the CRLB in this regime for small values of N. However, it remains true that dataset (b) always provides more accurate MLE results than dataset (c), both in terms of bias and variance.

We also remark that in several cases, a lower MSE is obtained from assuming that σ is unknown rather than known. Such an empirical observation is rather counterintuitive, but it is not in contradiction with statistical theory. Indeed, we can observe that a lower MSE value comes with a larger bias magnitude, which is typical of the usual bias-variance tradeoff.

D. Extension to phased array magnitude MR images

An interesting extension is to consider the case of data following a generalized Rician distribution of positive integer degree. Such data can be found in phased array magnitude MR images, where multiple-receiver coils are employed, when sum-of-squares images are considered [7, 26, 27]. It is also the case in phase contrast MR images [26, 28, 29]. From C iid couples of complex data (M_{c1}, M_{c2}) defined by

$$M_{c1} = \eta_c \cos\phi + X_{c1}, \quad M_{c2} = \eta_c \sin\phi + X_{c2},$$

where $(X_{11}, X_{12}, \ldots, X_{C1}, X_{C2})$ are iid centered real-valued Gaussian variables of standard deviation $\alpha > 0$, one obtains that $X = \sqrt{\sum_{c=1}^{C}(M_{c1}^2 + M_{c2}^2)}$ is a generalized Rician variable of parameters (η, α) and of degree $q = C - 1$, with $\eta = \sqrt{\sum_{c=1}^{C}\eta_c^2}$. The corresponding probability density is given by (29).

Most results presented in the single receiver coil case can be extended to this generalized context. In particular, we have $\mathcal{I}_\eta = \Psi_q(\rho)/\alpha^2$, with

$$\Psi_q(\rho) = \int_0^\infty \frac{y^{q+3}}{\rho^{q+2}} \exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right) \frac{I_{q+1}^2(y)}{I_q(y)}\,\mathrm{d}y - \rho.$$

Subsection IV-C shows that Ψ_q remains an increasing function between 0 and 1 for all positive values of q, so that Proposition 1 remains valid. More specifically, for each voxel, the magnitude of the average of the collected data over N repetitions

$$\left(\sum_{c=1}^{C}\left(\sum_{n=1}^{N} M_{c1,n}^2 + \sum_{n=1}^{N} M_{c2,n}^2\right)\right)^{1/2}$$

contains more information regarding the intensity parameter η, than the set of N magnitudes

$$\left(\sum_{c=1}^{C}\left(M_{c1,n}^2 + M_{c2,n}^2\right)\right)^{1/2},$$

when the noise level is known. The case where the noise level is unknown remains to be explored, but it is fairly conceivable that Proposition 2 also admit a direct extension.

As concerns MLE in the multiple-receiver coil case, the expression of the NLL of dataset (c) becomes

$$J(\boldsymbol{A}, \sigma) = \sum_\ell J_\ell(A_\ell, \sigma) - \log \sigma^{2LN(q-1)} \tag{24}$$

with

$$J_\ell(A_\ell, \sigma) = \frac{NA_\ell^2}{2\sigma^2} + \log A_\ell^{Nq} + \sum_n \left(\frac{Y_{\ell n}^2}{2\sigma^2} - \log I_q\left(\frac{A_\ell Y_{\ell n}}{\sigma^2}\right) + \log Y_{\ell n}^{q-1}\right). \tag{25}$$

where $Y_{\ell n}$ combines the data acquired at all coils at voxel ℓ for the nth repetition. The MM algorithm proposed in Subsection III-C admits a natural extension, based on the fact that $x^{-q} I_q(x)$ is strictly log-convex for all $q > -1/2$ [20]. More specifically, the update equation (21) becomes

$$A_{\ell k+1} = \frac{1}{N} \sum_n r_q\left(\frac{A_{\ell k} Y_n}{\sigma^2}\right) Y_n, \tag{26}$$

with a preserved non-decreasing effect on the likelihood. Moreover, Proposition 3 can be extended to the NLL (25) since noncentral chi-square distributions of more than two degrees of freedom have a strictly log-concave density (if their noncentrality parameter is positive) [18]. Therefore, the NLL (24) is a unimodal function of A_ℓ.

The numerical study proposed in Subsection III-C should be extended to check if the performance of MLE in terms of bias and variance remains in conformity with the analysis based on Fisher information.

IV. Mathematical analysis

A. Estimation of η when α is known

The Fisher information $\mathcal{I}_\eta$ is expressed through (4). Unfortunately, the remaining integral in (4) does not have a closed form, which complicates the analytical study of $\mathcal{I}_\eta$. In order to overcome the obstacle, the present contribution concentrates on the univariate function Ψ defined by (6). Firstly, lower and upper bounds are exhibited for Ψ. Then, a refined study allows us to conclude that Ψ is an increasing function of parameter $\rho = \eta^2/\alpha^2$.

Theorem 1 *Function Ψ is bounded below and above according to*

$$\max\left\{\Psi_{\mathrm{L1}}(\rho), \Psi_{\mathrm{L2}}(\rho)\right\} \leqslant \Psi \leqslant \Psi_{\mathrm{U}}(\rho) \tag{27}$$

where

$$\begin{aligned} \Psi_{\mathrm{L1}}(\rho) &= \frac{\rho}{\rho+1}, \\ \Psi_{\mathrm{L2}}(\rho) &= \frac{\pi\rho\left(I_0\left(\frac{\rho}{4}\right)+I_1\left(\frac{\rho}{4}\right)\right)^2}{8(2+\rho)e^{\frac{\rho}{2}}-\pi\left((2+\rho)I_0\left(\frac{\rho}{4}\right)+\rho I_1\left(\frac{\rho}{4}\right)\right)^2} \\ \Psi_{\mathrm{U}}(\rho) &= \min\left\{\rho, 1\right\}. \end{aligned} \tag{28}$$

Proof: See Appendix A. ■

Fig. 2 illustrates the behavior of function Ψ. Let us remark that Ψ_{L2} provides a very tight approximation of Ψ. It can be numerically checked that it is larger than Ψ_{L1} except for small values of ρ (the crossing between Ψ_{L1} and Ψ_{L2} is around $\rho_0 \simeq 0.6$).

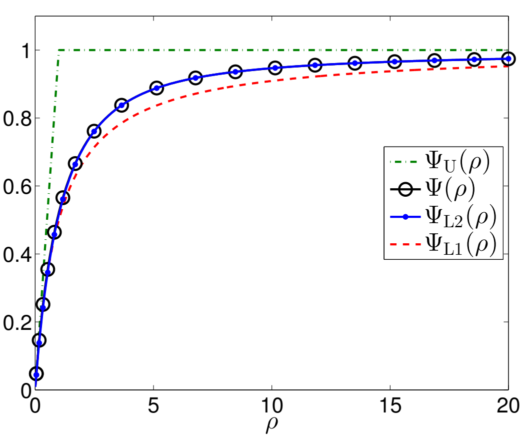

Fig. 2. Behavior of function $\Psi(\rho) = \alpha^2 \mathcal{I}_\eta$ and of its bounds according to Theorem 1.

In what follows, we manipulate statistical averages with respect to a so-called generalized Rician law of degree 1. More generally, let us define the generalized Rician law of degree q by its density

$$R_q(x\ ;\ \eta, \alpha) = \frac{x^{q+1}}{\alpha^2\eta^q}\exp\left(-\frac{x^2+\eta^2}{2\alpha^2}\right) I_q\left(\frac{\eta x}{\alpha^2}\right) \tag{29}$$

over $\mathbb{R}^+$, with $\eta \geqslant 0$, $\alpha > 0$, and the range of q being extensible to $(-1, \infty)$. The same probability law is also known as the noncentral chi distribution of degree $\nu = 2(q+1)$ when $\alpha = 1$ [15]. Moreover, let $\mathrm{E}_q[\cdot]$ denote the expectation operator relative to the generalized Rician law of degree q, and let $r_q = I_{q+1}/I_q$.

Proposition 5 *The Fisher information $\mathcal{I}_\eta$ admits the following expression:*

$$\mathcal{I}_\eta = \mathrm{E}_1\left[\frac{\eta X}{\alpha^4}\left(r_0\left(\frac{\eta X}{\alpha^2}\right) - r_1\left(\frac{\eta X}{\alpha^2}\right)\right)\right]. \tag{30}$$

Proof: The first part of the expectation is an obvious, alternate expression for the integral term in (4). On the other hand, we have

$$\mathrm{E}_1\left[\frac{\eta X}{\alpha^4} r_1\left(\frac{\eta X}{\alpha^2}\right)\right] = \frac{1}{\alpha^6}\int_0^\infty x^3 \exp\left(-\frac{x^2+\eta^2}{2\alpha^2}\right) I_2\left(\frac{\eta x}{\alpha^2}\right) \mathrm{d}x = \frac{\eta^2}{\alpha^4}$$

given general integration results related to the confluent hypergeometric function of the first kind [24]. Thus, expression (30) identifies with (4). ■

Let us remark that the Turán-type inequality $I_q^2(x) > I_{q-1}(x) I_{q+1}(x)$ for all $x > 0$, $q > -1$ [30–32], also reads $r_{q-1}(x) > r_q(x)$, which makes it apparent that $\mathcal{I}_\eta \geqslant 0$. Of course, the nonnegativity of the Fisher information holds by definition, but other Turán-type inequalities will be helpful to study additional properties of $\mathcal{I}_\eta$.

In (30), the expectation is relative to the density $R_1(\cdot\ ;\eta,\alpha)$. As a consequence, $Y = \eta X/\alpha^2$ is distributed according to the single parameter density

$$R_1(y\ ;\rho,\sqrt{\rho}) = \frac{y^2}{\rho^2}\exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right) I_1(y), \tag{31}$$

where $\rho = \eta^2/\alpha^2$. Therefore, according to (30), we have $\Psi(\rho) = \alpha^2 \mathcal{I}_\eta = \mathrm{E}[\psi(Y)]$ with

$$\psi(y) = y\,(r_0(y) - r_1(y)). \tag{32}$$

This is a key expression to derive the following theorem.

Theorem 2 *Function Ψ is increasing and is such that*

$$\lim_{\rho\to 0}\Psi(\rho) = 0, \quad \lim_{\rho\to 0}\Psi'(\rho) = 1, \quad \lim_{\rho\to\infty}\Psi(\rho) = 1. \tag{33}$$

Proof: See Appendix B. ■

It is worth mentioning that the proof of Proposition 2 manipulates stochastic order properties to establish a monotonicity property of an information metric, akin to [33] in the context of performance evaluation of communication channels. The latter paper considers the case of Rician fading channels, where the received signal amplitude follows a Rician distribution. However, Proposition 2 is not a consequence of results contained in [33], and the proof of Proposition 2 is not adapted from [33].

Let us remark that the increasing character of Ψ implies that $\Psi(\rho)$ is strictly lower than 1 for all ρ. Let us also mention that direct differentiation under the integral sign in (6) provides

$$\Psi'(\rho) = \frac{1}{2\rho^4}\int_0^\infty y^5 \exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right)\frac{I_1^2(y)}{I_0(y)}\,\mathrm{d}y - \frac{4+\rho}{2\rho^3}\int_0^\infty y^3\exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right)\frac{I_1^2(y)}{I_0(y)}\,\mathrm{d}y - 1.$$

Given $\mathrm{E}\left[Y^2\right] = \rho(\rho+4)$, the following remarkable identity can be obtained:

$$\Psi'(\rho) = \frac{1}{2\rho^2}\left(\mathrm{E}\left[Y^2\psi(Y)\right] - \mathrm{E}\left[Y^2\right]\mathrm{E}\left[\psi(Y)\right]\right). \tag{34}$$

Therefore, Theorem 2 expresses that the correlation between variables Y^2 and $\psi(Y)$ is positive.

B. Estimation of η_ℓ when α is unknown in the case of several Rician variables

The purpose of the present subsection is to examine some mathematical properties of function $\tilde{\Psi}_\ell$ defined by (11), and more specifically the variations of $\tilde{\Psi}_\ell(\mu\boldsymbol{\rho})$ as a function of $\mu \geqslant 0$.

Proposition 6 *For any intensity vector $\boldsymbol{\rho}$ and any voxel ℓ,*

$$\Psi(\rho_\ell) \geqslant \tilde{\Psi}_\ell(\boldsymbol{\rho}) \geqslant \tilde{\Psi}(\rho_\ell), \tag{35}$$

$$\lim_{\mu\to 0} \tilde{\Psi}_\ell(\mu\boldsymbol{\rho}) = 0, \tag{36}$$

$$\lim_{\mu\to\infty} \tilde{\Psi}_\ell(\mu\boldsymbol{\rho}) = 1. \tag{37}$$

Moreover, $\tilde{\Psi}_\ell(\mu\boldsymbol{\rho})$ is an increasing function of μ if inequality (18) *holds.*

Proof: See Appendix C ∎

Numerical evaluations indicate that inequality (18) holds (see Fig. 3 for an illustration). Unfortunately, we have been unable to prove it mathematically. Let us remark that it corresponds to a stronger condition than the increasing character of Ψ, which is the main result of Theorem 2.

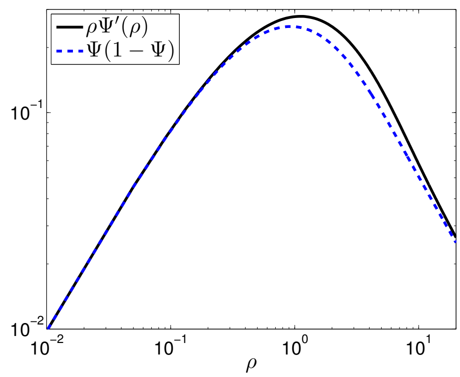

Fig. 3. Numerical evaluation of (18), in log-log scale.

C. Extension to the generalized Rician case

Theorem 1, Proposition 5 and Theorem 2 naturally extend to positive values of q. In particular, we have $\Psi_q(\rho) = \mathrm{E}[\psi_q(Y)]$, where Y is distributed according to $R_{q+1}(y\ ;\ \rho, \sqrt{\rho})$, and $\psi_q(y) = y\,(r_q(y) - r_{q+1}(y))$. Inequalities (39) allow us to conclude that ψ_q and Ψ_q take values between 0 and 1 for all $q \geqslant 0$. Moreover, along the same lines as the proof of Proposition 5, we obtain that $\Psi_{q\mathrm{L}}(\rho) \leqslant \Psi_q(y) \leqslant \Psi_{q\mathrm{U}}(\rho)$ with

$$\Psi_{q\mathrm{L}}(\rho) = \frac{\rho}{\rho+q+1}, \qquad \Psi_{q\mathrm{U}}(\rho) = \min\left\{\frac{\rho}{q+1}, 1\right\}.$$

Let us indicate that $\Psi_{q\mathrm{L}} = \Psi_{\mathrm{L}1}$. It also seems possible to derive an generalized expression of $\Psi_{\mathrm{L}2}$, but we have not pursued in this direction.

On the other hand, Theorem 2 has a direct extension. In particular, Ψ_q is an increasing function for all $q \geqslant 0$, the proof of Appendix B being directly applicable. Moreover, (34) admits the following generalization:

$$\Psi_q'(\rho) = \frac{1}{2\rho^2}\left(\mathrm{E}\left[Y^2\psi_q(Y)\right] - \mathrm{E}\left[Y^2\right]\mathrm{E}\left[\psi_q(Y)\right]\right).$$

We can thus conclude that Y^2 and $\psi_q(Y)$ are positively correlated. Finally, we have the following proposition.

Proposition 7 $\Psi_q(x)$ *is a nonincreasing function of q for all $x > 0$, $q > -1$.*

Proof: Let $\chi_{\nu,\eta}$ denote the noncentral chi distribution of noncentrality parameter η and of degree ν. Function $\Psi_q(\rho)$ is the Fisher information of a variable $X \sim \chi_{\nu,\eta}$ with $\nu = 2(q+1)$ and $\eta = \sqrt{\rho}$, related to the estimation of η. Noncentral chi-squared distributions are reproductive under convolution [15], so that $X' = \sqrt{X^2 + D^2} \sim \chi_{\nu+d,\eta}$ if $D \sim \chi_{d,0}$, $d > 0$, D being independent from X. Consequently, the Fisher information of X' related to η is $\Psi_{q'}(\rho)$ with $q' = q + d/2$. According to the data-processing inequality [14], the Fisher information of X' is not larger than the Fisher information of X, since D conveys no information about η, which allows to conclude. ∎

Fig. 4 illustrates the nonincreasing character of $\Psi_q(x)$ as a function of q.

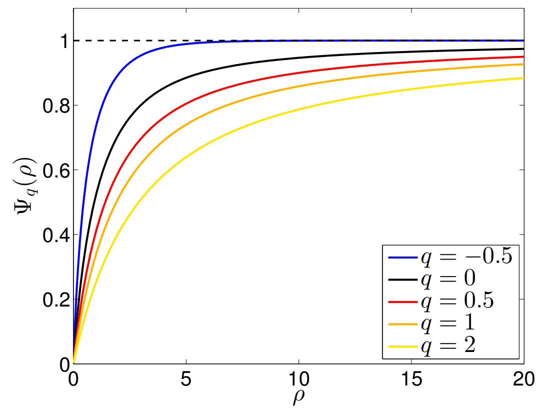

Fig. 4. Numerical evaluation of $\Psi_q(x)$ for several values of $q > -1$. Let us remind that $\Psi = \Psi_0$.

V. Conclusion

When several repeated MRI acquisitions are performed, the standard procedure consists in relying on the magnitude of the mean of the complex data available at each voxel. Our study leads to the conclusion that such a preprocessing step is preferable to considering the magnitude of each data separately, from the information theoretic standpoint. The key mathematical result underlying our study is the fact that, for Rician distributions, the Fisher information with respect to the noncentrality parameter is an increasing function, at constant noise level. We have obtained such a result by combining stochastic order properties of Rician distributions, with inequalities involving modified Bessel functions. Let us stress that the exact evaluation of the Fisher information is impossible in this context, since its expression involves modified Bessel functions within an integral term that can only be approximated numerically. Previous studies devoted to MRI such as [2, 4, 16] already displayed (numerical approximations of) the Fisher information, and empirically noticed or implicitely assumed its monotonous behavior. The present paper provides a rigorous validation of such a behavior. Moreover, the validity of the latter is extended to the case of multiple-coil MR acquisition systems coupled with sum-of-squares reconstruction. We have also studied the single-coil case when the noise level is unknown, with essentially the same conclusion, although our mathematical treatment is not fully complete in this case.

A significant part of the paper has focused on the maximum likelihood solution. In particuler, we have empirically checked that the effect of the bias at low signal to noise ratio does not modify the previous conclusion. We have also proposed an efficient way of computing the maximum likelihood estimator, based on a majorization-minimization algorithm, and we have shown that it remains applicable in a wide range of situations.

Finally, let us stress that our study is devoted to an MRI problem, but its information theoretic content may be also of interest in several engineering domains where Rician distributions also play an important role, such as ultrasound imaging, radar signal processing, and digital communications.

Acknowledgment

The first author is indebted to Árpad Baricz for several useful comments and bibliographical suggestions.

Appendix

A. Proof of Theorem 1

The first lower bound $\Psi_{\mathrm{L}1}(\rho) \leqslant \Psi$ is a direct consequence of the Cramér-Rao inequality [11, 12.333]

$$\operatorname{var}(T) \geqslant \frac{1}{\mathcal{I}_\eta}\left(\frac{\partial \mathrm{E}\left[T\right]}{\partial \eta}\right)^2 \tag{38}$$

applied to the biased estimator $T = X^2$. Indeed, we have $\partial \mathrm{E}\left[T\right]/\partial\eta = 2\eta$ since $\mathrm{E}\left[X^2\right] = 2\alpha^2 + \eta^2$, and also $\mathrm{E}\left[T^2\right] = \mathrm{E}\left[X^4\right] = 8\alpha^4 + 8\alpha^2\eta^2 + \eta^4$, thus $\operatorname{var}(T) = 4\alpha^4 + 4\alpha^2\eta^2$. The expected result then follows from (38).

In the same manner, the second lower bound $\Psi_{\mathrm{L}2}(\rho) \leqslant \Psi$ derives from the application of (38) to $T = X$. The obtained expression is more complex, resulting from the fact that the expectation of the Rician law reads [4, 26, 28]:

$$\mathrm{E}[X] = \sqrt{2\pi}\frac{\alpha}{4}e^{-\rho/4}\left((2+\rho)I_0\left(\frac{\rho}{4}\right) + \rho I_1\left(\frac{\rho}{4}\right)\right).$$

The calculation of $\Psi_{\mathrm{L}2}$ makes use of the identities $I_0'(x) = I_1(x)$ and $I_1'(x) = I_0(x) - I_1(x)/x$.

The fact that $\Psi \leqslant 1$ can be derived using the data-processing inequality [11, 14], which states that data processing cannot increase the information content of a given dataset. On the one hand, (3) expresses the Fisher information related to η carried by the Gaussian couple (Y_1, Y_2). On the other hand, the data-processing inequality yields $\mathcal{I}_\eta(Y_1^2 + Y_2^2) \leqslant \mathcal{I}_\eta(Y_1, Y_2)$, which allows to conclude since $X = \sqrt{Y_1^2 + Y_2^2}$.

Finally, $\Psi \leqslant \rho$ results from the Turán-type inequality [30, 32, 34] (which holds for all $q > -1, y > 0$)

$$I_q^2(y) - I_{q-1}(y)I_{q+1}(y) \leqslant \frac{1}{q+1}I_q^2(y),$$

applied to $q = 1$, *i.e.*, $I_1^2(y) \leqslant 2I_0(y)I_2(y)$ for all $y > 0$. Given (6), the expected upper bound is readily obtained from the fact that

$$\int_0^\infty \frac{y^3}{\rho^2}\exp\left(-\frac{y^2}{2\rho} - \frac{\rho}{2}\right) I_2(y)\,\mathrm{d}y = \rho.$$

B. Proof of Theorem 2

Let us observe that the values of ψ lie between 0 and 1, according to the inequalities [31]

$$0 < r_q(y) - r_{q+1}(y) < \frac{1}{y} \tag{39}$$

where $q \geqslant 0$ and $y > 0$. We thus obtain that $\Psi(\rho)$ lies also between 0 and 1, which we already known according to Theorem 1.

The limit values (33) are immediate consequences of the bounds of Theorem 1. It remains to show that function Ψ is increasing. To derive such a property, we will establish that

(i) the single parameter density $p_\rho = R_1(\cdot\,;\,\rho, \sqrt{\rho})$ defines a univariate *stochastically ordered family* [35, 36], for which the stochastic ordering is strict in the sense of [37].

(ii) function ψ defined by (32) is increasing.

The conclusion will then be a direct consequence of [37, Propositions A.1 and A.2].

Given two real valued random variables X, X', X is smaller than X' in the usual stochastic order if the following inequality holds between their cumulative distributions:

$$P(X \leqslant x) \geqslant P(X' \leqslant x) \tag{40}$$

for all x. Moreover, X is strictly stochastically smaller than X' if (40) holds, and $P(X \leqslant x^*) < P(X' \leqslant x^*)$ for some x^* [37]. In order to prove statement (i), let us first remark that the cumulative distribution of Y can be put into the following form: $P(Y \leqslant y) = 1 - Q_2\left(\sqrt{\rho}, y/\sqrt{\rho}\right)$, where $Q_\nu(a, b)$ is the generalized Marcum Q-function. Statement (i) is then easily deduced from known properties of $Q_\nu(a, b)$, and more specifically from the fact that

$Q_\nu(a,b)$ is strictly increasing in a for all $a \geqslant 0$ and b, $\nu > 0$ and strictly decreasing for all a, $b \geqslant 0$ and $\nu > 0$ [38, Theorem 1(a)].

The proof of statement (ii) is less straightforward. On the one hand, according to [32, 34],

$$(yr_q(y) - yr_{q+1}(y))' = y\frac{r_{q-1}(y)\, r_{q+1}(y) - r_q^2(y)}{r_{q-1}(y)\, r_q(y)}$$

for all q and all $y > 0$. On the other hand, $r_\nu(x)$ is a log-convex function of ν, for all $\nu > -1$ and all $x > 0$ [39], which implies the Turán type inequality

$$r_{q-1}(y)\, r_{q+1}(y) - r_q^2(y) > 0 \tag{41}$$

for all $q > 0$ and all $y > 0$. The result we need is (41) for $q = 0$. However, according to the Gale-Klee-Rockafellar theorem [40], the log-convexity of $r_\nu(x)$ also holds at $\nu = -1$ since $r_\nu(x)$ is a continuous function of ν, which allows us to conclude that (41) holds[1] for $q = 0$ and $y > 0$ (which is equivalent to $r_1(y) > r_0^3(y)$ for all $y > 0$, *i.e.*, $I_0^3(y)I_2(y) > I_1^4(y)$). Hence, $\psi(y) = y\,(r_0(y) - r_1(y))$ is increasing for all $y > 0$.

C. Proof of Proposition 6

Let us first remark that f_ℓ defined by (12) also reads $f_\ell(\boldsymbol{\rho}) = 1 + \sum_{i\neq\ell}(\rho_i + 1)\,(1 - \Psi_{\mathrm{L1}}(\rho_i)/\Psi(\rho_i))$ where Ψ_{L1} is defined by (28). Since $\Psi \geqslant \Psi_{\mathrm{L1}} \geqslant 0$, we have $f_\ell(\boldsymbol{\rho}) \geqslant 1$, so that $\tilde{\Psi}_\ell(\boldsymbol{\rho}) \geqslant \tilde{\Psi}(\rho_\ell)$ for all $\boldsymbol{\rho}$. We have also $D_\ell(\boldsymbol{\rho}) > 0$ for all $\boldsymbol{\rho}$, so we can deduce $\Psi(\rho_\ell) \geqslant \tilde{\Psi}_\ell(\boldsymbol{\rho})$ from (11) (the inequality being strict for all $\rho_\ell > 0$ since $\Psi < 1$ according to Theorem 2).

Since $\Psi(\rho_\ell) \geqslant \tilde{\Psi}_\ell(\boldsymbol{\rho}) \geqslant 0$ and $\lim_{\rho\to 0}\Psi(\rho) = 0$, it is immediate to obtain (36). On the other hand, we have $1 > \Psi(\rho_\ell) \geqslant \tilde{\Psi}_\ell(\boldsymbol{\rho}) \geqslant \tilde{\Psi}(\rho_\ell)$ for all $\boldsymbol{\rho}$. In order to conclude that (37) holds, let us prove $\lim_{\rho\to\infty}\tilde{\Psi}(\rho) = 1$. According to Theorem 1, Ψ_{L2} is a lower bound of Ψ. Moreover, an asymptotic expansion of Ψ_{L2} for large values of ρ gives $\Psi_{\mathrm{L2}}(\rho) = 1 - 1/2\rho + o\,(1/\rho)$, and hence, $\lim_{\rho\to\infty}(\rho\Psi_{\mathrm{L2}} - \rho + 1) = 1/2$. Therefore, for any small positive constant c and for sufficiently large values of ρ, we have

$$\tilde{\Psi}(\rho) = 1 - \frac{1 - \Psi(\rho)}{\rho\Psi(\rho) - \rho + 1} \geqslant 2\Psi(\rho) - 1 - c,$$

which allows to conclude that (37) holds by considering that the value of c can be chosen arbitrary small.

Finally, let us show that (18) implies that $\tilde{\Psi}_\ell(\boldsymbol{\rho})$ is coordinatewise increasing, and hence that the last assertion of Proposition 6 is true. For all $j \neq \ell$, given (11), it suffices to show that $f_\ell(\boldsymbol{\rho})$ is an increasing function of ρ_j. The conclusion is immediate since

$$\frac{\partial f_\ell(\boldsymbol{\rho})}{\partial \rho_j} = \frac{\rho_j\Psi'(\rho_j) - \Psi(\rho_j)(1 - \Psi(\rho_j))}{\Psi^2(\rho_j)}.$$

It remains to show that $\tilde{\Psi}_\ell(\boldsymbol{\rho})$ is also an increasing function of ρ_ℓ if (18) holds. Straightforward calculations provide

$$\frac{\partial f_\ell(\boldsymbol{\rho})}{\partial \rho_\ell} = \frac{f_\ell(\boldsymbol{\rho})}{D_\ell^2(\boldsymbol{\rho})}\left(f_\ell(\boldsymbol{\rho})\Psi'(\rho_\ell) - (1 - \Psi(\rho_\ell))^2\right),$$

which is positive for all $\boldsymbol{\rho}$ since $f_\ell \geqslant 1$ and $\Psi(\rho)(1 - \Psi(\rho)) \geqslant \rho(1 - \Psi(\rho))^2$ for all $\rho \geqslant 0$.

[1] Rigorously, we first conclude that inequality (41) holds in the wide sense for $q = 0$ and all $y > 0$. The strict inequality can then be checked by contradiction: for some $x > 0$, assume that $r_{-1}(x)\, r_1(x) = r_0^2(x)$. $r_\nu(x)$ would then be a linear function of ν for $\nu \in (-1, 1]$, a contradiction with the strict log-convexity of $r_\nu(x)$ in $(-1, 1]$.

REFERENCES

[1] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data", *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, 1995.
[2] J. Sijbers and A. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance from MR data", *Magn. Reson. Med.*, vol. 51, no. 3, pp. 586–594, 2004.
[3] T. Benedict and T. Soong, "The joint estimation of signal and noise from the sum envelope", *IEEE Trans. Inf. Theory*, vol. 13, pp. 447–454, 1967.
[4] O. Karlsen, R. Verhagen, and W. Bovée, "Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: application to T1 and perfusion measurements", *Magn. Reson. Med.*, vol. 41, no. 3, pp. 614–623, 1999.
[5] J. L. Andersson, "Maximum a posteriori estimation of diffusion tensor parameters using a Rician noise model: Why, how and but", *Neuroimage*, vol. 42, no. 4, pp. 1340–1356, 2008.
[6] D. C. Alexander, "Modelling, fitting and sampling in diffusion MRI", in *Visualization and Processing of Tensor Fields*, D. Laidlaw and J. Weickert, Eds. Springer Verlag, 2009.
[7] C. D. Constantinides, E. Atalar, and E. R. McVeigh, "Signal-to-noise measurements in magnitude images from NMR phased arrays", *Magn. Reson. Med.*, vol. 38, no. 5, pp. 852–857, 1997.
[8] S. Rice, "Mathematical analysis of random noise", *Bell Syst. Tech. J.*, vol. 24, pp. 46–156, 1945.
[9] H. L. Van Trees, *Detection, estimation, and modulation theory, part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*, John Wiley, New York, NY, 2004.
[10] J. G. Proakis, *Digital Communications*, McGraw-Hill, Singapore, 2001.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley and Sons, New York, NY, 1991.
[12] D. H. J. Poot, "Advances in the reconstruction and statistical processing of magnetic resonance images", PhD thesis, University of Antwerp, Belgium, 2010.
[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ. Press, New York, 3rd edition, 1992.
[14] R. Zamir, "A proof of the Fisher information inequality via a data processing argument", *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246–1250, 1998.
[15] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 2, Wiley, 2nd edition, 1995.
[16] M. D. Tisdall, R. A. Lockhart, and M. S. Atkins, "The bias/variance trade-off when estimating the MR signal magnitude from the complex average of repeated measurements", *Magn. Reson. Med.*, vol. 66, no. 5, pp. 1456–1467, 2011.
[17] J. Sijbers, A. den Dekker, P. Scheunders, and D. V. Dyck, "Maximum likelihood estimation of Rician distribution parameters", *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357–361, 1998.
[18] S. Das Gupta and S. K. Sarkar, *On* $\mathrm{TP}\boldsymbol{s}2$ *and log-concavity*, vol. 5 of *Lecture Notes-Monograph Series*, pp. 54–58, Institute of Mathematical Statistics, Hayward, CA, 1984.
[19] A. V. Boyd, "The monotonicity of $I_1(x)/I_0(x)$", *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz.*, vol. 209, pp. 35–36, 1967.
[20] E. Neuman, "Inequalities involving modified Bessel functions of the first kind", *J. Math. Anal. Appl.*, vol. 171, no. 2, pp. 532 – 536, 1992.
[21] D. Varadarajan and J. Haldar, "A quadratic majorize-minimize framework for statistical estimation with noisy Rician and noncentral chi-distributed MR images", in *Proc. IEEE ISBI*, Apr. 2013, pp. 712–715.
[22] D. R. Hunter and K. Lange, "A tutorial on MM algorithms", *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.
[23] V. Solo and J. Noh, "An EM algorithm for Rician fMRI activation detection", in *Proc. IEEE ISBI*, Apr. 2007, pp. 464–467.
[24] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*, Dover publications, New York, NY, 1970.
[25] M. Jacobson and J. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms", *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, Oct. 2007.
[26] J. Sijbers, *Estimation of Signal and Noise from Magnitude Magnetic Resonance Images*, PhD thesis, University of Antwerp, Belgium, 1998.
[27] C. G. Koay and P. J. Basser, "Analytically exact correction scheme for signal extraction from noisy magnitude MR signals", *J. Magn. Reson.*, vol. 179, no. 2, pp. 317–322, 2006.
[28] A. H. Andersen and J. E. Kirsch, "Analysis of noise in phase contrast MR imaging", *Math. Program.*, vol. 23, no. 6, pp. 857–869, 1996.
[29] J. Sijbers, A. den Dekker, M. Verhoye, and D. V. Dyck, "Optimal estimation of T2 maps from magnitude MR data", in *Proc. SPIE Medical Imag.*, K. M. Hanson, Ed., San Diego, CA, USA, Feb. 1998, vol. 3338, pp. 384–390.
[30] V. Thiruvenkatachar and T. Nanjundiah, "Inequalities concerning Bessel functions and orthogonal polynomials", *Proc. Indian Acad. Sci. - Section A*, vol. 33, no. 6, pp. 373–384, 1951.
[31] D. E. Amos, "Computation of modified Bessel functions and their ratios", *Mathematics of Computation*, vol. 28, no. 125, pp. 239–251, 1974.
[32] A. Baricz, "Turán type inequalities for modified Bessel functions", *Bulletin of the Australian Mathematical Society*, vol. 82, pp. 254–264, Oct. 2010.
[33] C. Tepedelenlioglu, A. Rajan, and Y. Zhang, "Applications of stochastic ordering to wireless communications", *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4249–4257, 2011.
[34] C. M. Joshi and S. K. Bissu, "Some inequalities of Bessel and modified Bessel functions", *J. Austral. Math. Soc. (Series A)*, vol. 50, pp. 333–342, Apr. 1991.
[35] E. L. Lehmann, "Ordered families of distributions", *Ann. Math. Statist.*, vol. 26, pp. 399–419, 1955.
[36] M. Shaked and J. G. Shanthikumar, *Stochastic orders*, Springer, 2007.
[37] E. Altman, G. Neglia, F. De Pellegrini, and D. Miorandi, "Decentralized stochastic control of delay tolerant networks", Tech. Rep. RR-6654, INRIA, 2008.

[38] Y. Sun, A. Baricz, and S. Zhou, "On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q-functions", *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1166–1186, Mar. 2010.
[39] A. Baricz, "Bounds for modified Bessel functions of the first and second kinds", *Proc. Edinburgh Math. Soc. (Series 2)*, vol. 53, pp. 575–599, Oct. 2010.
[40] D. Gale, V. Klee, and R. Rockafellar, "Convex functions on convex polytopes", *Proc. Amer. Math. Soc.*, vol. 19, pp. 867–873, 1968.